DOCUMENT RESUME

ED 283 833 TM 870 322

AUTHOR Marso, Ronald N.; Pigge, Fred L.
TITLE A State-Wide Assessment of the Testing and Evaluation Needs and Proficiencies of Beginning Teachers: Implications for Staff Development.
PUB DATE Mar 87
NOTE 30p.; Paper presented at the Annual Meeting of the Association of Supervision and Curriculum Development (New Orleans, LA, March 21-24, 1987).
PUB TYPE Speeches/Conference Papers (150) -- Reports - Research/Technical (143)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Administrator Attitudes; *Beginning Teachers; Elementary School Teachers; Elementary Secondary Education; *Needs Assessment; Rating Scales; Secondary School Teachers; State Surveys; *Teacher Evaluation; *Teacher Made Tests; *Teaching Skills; Test Construction; Training Objectives
IDENTIFIERS Ohio

ABSTRACT

In order to further existing knowledge about classroom testing practices, this study surveyed Ohio supervisors' and principals' perceptions of beginning teachers' needs for and proficiency in selected testing and evaluation competencies. Respondents considered typical beginning elementary and secondary teachers, rather than teachers of special education, music, art, or physical education. Usable survey responses to the 26-item rating scale were returned by 229 teacher supervisors, 313 building principals, and 44 curriculum or instruction coordinators. Four hypotheses were examined. First, beginning teachers' needs and proficiencies are similar, indicating they are well trained in testing and evaluation competencies, specifically test construction and test score use. Second, levels of training are equal for different grade levels and for rural, urban, or suburban schools. Third, testing competencies are as high or higher than teachers' reported subject knowledge, professional education competencies, and overall competencies. Fourth, supervisors and principals will agree on beginning teachers' needs, as well as on their proficiencies. Results indicated the following: (1) the teachers' proficiencies were inadequate to meet the job needs; (2) hypothesis 2 was supported; (3) hypothesis 3 was rejected; and (4) principals and supervisors agreed on teachers evaluation competency needs but disagreed on proficiencies. (GDC)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED283833

A State-Wide Assessment of the Testing and Evaluation Needs and Proficiencies of Beginning Teachers: Implications for Staff Development

Ronald N. Marso

Fred L. Pigge

College of Education and Allied Professions

Bowling Green State University

Bowling Green, Ohio 43403

A paper presented at the annual meeting of the

Association of Supervision and Curriculum Development

New Orleans, LA

March 21-24, 1987

Running head: STATE-WIDE ASSESSMENT

TM 870 322

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

R. Marso

F. Pigge

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☒ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

BEST COPY AVAILAB

A State-Wide Assessment of the Testing and Evaluation Needs and Proficiencies of Beginning Teachers: Implications for Staff Development

Despite the fact that teacher-made tests occupy far more classroom time and effort than do standardized tests, standardized tests have won the attention and interest of researchers, media, and the general public (Coffman, 1971; Fleming & Chambers, 1983). However, the interests of the general public in standardized testing can best be described as running sweet and sour much to the chagrin of the measurement profession. The current strident demands for increased educational accountability and higher educational standards via testing have followed so closely on the heels of demands for standardized testing moratoriums that one hears echos of cries of testing social injustice and discrimination concomitant with cries of elitism (Green, 1975; Madaus, 1985).

Nevertheless, public attention and research interest have facilitated a large and growing body of knowledge about standardized tests and testing; whereas, research on teacher-made tests and testing practices in the public school classrooms has largely been neglected. Further, the limited research on teacher-made tests and their use is restricted in scope by a preponderance of studies having been conducted in college classrooms and having been primarily limited to investigations of test reliability and test item characteristics. The gravity of this situation is such that

Dwyer (1982) maintained that the advice given to preservice and inservice teachers regarding the use of teacher-made tests in the public schools reflected a concensus of professional judgments rather than a foundation of empirical research. Similarly, Gullickson (1984) stated that we do not know whether classroom tests are used effectively or how they are used.

Some research that has been conducted in the public schools does provide a few suggestions about the effects of certain testing practices although we do not know whether or not teachers are using these practices. Further, these suggestions appear to be consistent with those suggestions derived from the earlier investigations conducted on college-age subjects (Balch, 1964). For example, recent research in the public school classrooms suggests: that students prepare differently for varied test item types (D'Ydewalle, Swerts, & DeCorte, 1983; Kulhavy, Dyer, & Silver, 1975), that students have preference for certain test item types (Shaha, 1984), that certain types of feedback following a test enhances learning and other types apparently do not (Hanna, 1976; Stewart & White, 1976; Wexley & Thornton, 1972), that the frequent administration of tests designed to facilitate learning tends to do so (Peckham & Roe, 1977), and that time spent in testing may be more efficient in promoting learning than comparable time spent in reviewing content (Nungester & Duchastel, 1982).

In reporting on one of the few investigations of testing attitudes and practices in the public school classrooms, Gullickson (1984) described the existing research on teacher-made testing practices as limited and idiosyncratic. She conducted a state-wide survey of third, seventh, and tenth-grade teachers regarding their attitudes toward the effects of classroom tests on student-learning, constraints associated with testing (e.g. availability of scoring assistance), effectiveness of tests as an evaluative tool (e.g. do tests facilitate instruction?), and testing practices and beliefs (e.g. students dislike taking tests). Most of the teachers in her sample reported having taken a college measurement class. She concluded that teachers are very supportive of the use of classroom tests, are comfortable with their knowledge about and use of teacher-made tests, feel that tests should be given frequently, and feel that tests are helpful in the instructional process but that they have limited evaluative usefulness. Lambert (1980-81) obtained a nationwide sample of opinions about standardized and teacher-made tests and testing practices in the public schools from chairpersons of state legislative committees, from principal officials in state teacher associations, and from the deans of the three largest teacher training institutions in each state. He found both agreement and divergent opinions within and between the three groups sampled. Widely divergent opinions were identified on matters such as whether or not teacher training institutions

should offer instruction in tests and measurements (e.g. one-third of the deans reported that their colleges did not offer such a course and had no intention of doing so), whether or not norm-referenced tests should be used for educational program evaluation, whether or not existing standardized tests are biased, and whether or not multiple-choice tests really assess the competence of classroom teachers. Contrarily, he found considerable agreement on opinions such as classroom teachers have a generally negative attitude about standardized tests, the importance of teachers producing superior classroom tests, and that teachers should know more about standardized tests and their use. Lambert's general conclusion was that there is an apparent need to make all three groups used in his nationwide survey more aware of both the values and limitations of tests.

A single study was identified which specifically addressed the overall nature of teacher-made testing practices in the public schools. Rogers (1985) had each of 89 university students registered for a tests and measurements class conduct an open-ended interview of an inservice public school teacher regarding the pupil evaluation process inclusive of test planning, test construction, testing for instruction, grading of students, and use of standardized tests. The inductive data analysis employed to interpret the reports of the university students

following the open-ended interviews resulted in the following generalizations: most of the teachers interviewed used paper and pencil tests, most used both self-constructed and publisher-made tests, most planned tests around curriculum guide objectives rather than test specification tables, most of the teachers used the percentage correct method of scoring, and the teachers varied considerably in the extent of their use of standardized test scores.

Purpose

The general purpose of this state-wide survey was to further existing knowledge about classroom testing practices through investigating supervisors' and principals' perceptions of beginning teachers' need for and proficiency in selected classroom testing and evaluation competency areas. The administrators were asked to use "typical" beginning elementary or secondary content teachers as frames-of-reference when completing the survey form. In other words, they were asked not to consider the testing and evaluation needs and proficiencies of beginning teachers in the special education and specialized (music, art, physical education) areas. Also, they were asked to consider the "typical" beginning elementary or content area teacher, not the best nor the poorest, just those teachers in the "middle" who might be classified as "typical."

The following four general hypotheses guided the investigation:

1) The supervisors and principals will report that beginning teachers are well trained relative to the competencies needed for classroom testing and evaluation. Specifically, the principals' and supervisors' ratings of the extent that competencies are needed will not differ significantly from their ratings of the proficiency of beginning teachers for: a) test development competencies, or b) test score use competencies.

2) The supervisors and principals with different grade level and school type assignments will report that beginning teachers are equally well trained in testing and evaluation competencies. Specifically, the ratings of beginning teachers' testing proficiencies and needs will not differ significantly when the rating supervisors or principals are assigned to: a) elementary as compared to middle or high school grades, and b) rural as compared to urban or suburban schools.

3) The supervisors and principals will report that beginning teachers' competencies in classroom testing and evaluation are equivalent to the level of their other professional competencies. Specifically, the supervisors and principals will rate the beginning teachers' testing related competencies as high or higher than: a) knowledge of their subject areas, b) their other professional education competencies, and c) their overall competencies as educators.

4) The supervisors and principals will be in close agreement about beginning teachers' testing and evaluation needs and proficiencies. Specifically, the supervisors' ratings of beginning teachers' competencies when compared to the principals' ratings will not differ significantly for: a) needs, or b) proficiencies.

Method

A survey instrument was constructed and sent during the winter of 1986 to a stratified random sample of supervisors of teachers and building principals in Ohio. The names and addresses of the subjects were selected from the State directory of schools. The type of school system (city, exempted village, or county local), the job assignment (principal or supervisor), and school grade level (elementary, middle, or secondary) classifications were used as strata in the random selection process. A total of 800 survey forms were mailed from which 586 (73%) usable survey responses were obtained after two follow-up contacts of nonrespondents. A total of 229 supervisors, 313 building principals, and 44 individuals in related supervisory roles (coordinators of curriculum or instruction, etc.) returned usable and completed survey forms.

The survey instrument consisted of a 26-item listing of competencies related to the development and use of teacher-made tests. These items were selected and reviewed for appropriateness by a team of five professors responsible for the instruction of

the tests and measurements course for preservice teachers at Bowling Green State University. The items then were grouped into two sections of the survey instrument with 17 items identified as test development competencies and nine items identified as test use competencies. Two five-point ('5' as high and '1' as low) Likert-type response scales were provided for each competency item and identified as: "need of this competency to be a successful teacher in your school" and "average proficiency of your new teachers in this competency." Each respondent was also asked to indicate the nature of his/her school(s) assignment (rural, urban, or suburban) and the grade level of his/her assignment (elementary, middle grades, secondary, K-12 grades, or other). Those respondents placing themselves in the "other" category were excluded from the analyses related to specific school assignments. Additionally, the respondents were asked to rate the preparation of their typical beginning teachers in tests and evaluation competencies via three Likert-type five-point scale items ('1' much below average to '5' well above average) relative to: the beginning teachers' subject area knowledge, the beginning teachers' knowledge and skill in other professional education competencies (planning, discipline, etc.), and the beginning teachers' overall (general) competencies as educators.

Results

Hypothesis One: Level of Needs Versus Proficiencies

A t test of the difference between dependent means was used to analyze the combined supervisor and principal ratings (N=586) of the beginning teachers' need for and proficiency in each identified competency. A t test was completed between the need mean and the proficiency mean for each of the 17 test development competency items and for each of the nine test score use competency items; also, t-ratios were completed on the totals for each of the two sections.

The t test analysis procedures resulted in the rejection of hypothesis one 'a' and 'b' as significant differences ($p < .001$) between the need and proficiency mean ratings were noted for each of the 26 competency items. Descriptions of the items, need and proficiency means, t-ratios, and other data related to these analyses are presented in Tables 1 and 2. For each of the 26 competency items the combined group of supervisors and principals rated the mean need for the competency significantly higher than they rated the typical beginning teachers' proficiency in that competency area. This would suggest that the supervisors and principals felt that the typical beginning teachers' test development and test score use proficiencies were inadequate to meet the needs of their jobs.

To better identify which beginning teacher competencies the respondents reported as being most deficient, a discrepancy index was calculated for each item (need mean minus proficiency mean) and each item was then ranked relative to this discrepancy index (see Table 1 and Table 2). The three items with the highest discrepancy "scores" for the test development competencies (see Table 1, items 9, 10, and 5b) were closely related to the impact of tests on pupil learning: writing questions demanding higher thinking processes, writing questions representing true student progress, and the scoring of essay questions. The three items with highest discrepancy "scores" for the test score use competencies (see Table 2, items 4, 9, and 5), similarly, were all associated with the use of tests to improve learning. Conversely, the items with lowest discrepancy scores on both sets of competency areas appeared to be skills (math calculations, grading, writing items, selecting items, or use of sociometric techniques) less directly related to the instructional-learning process.

Hypothesis Two: Needs and Proficiencies by Grade and School

A series of F tests were used to analyze the ratings of the principals and supervisors when classified by grade level assignment (elementary, middle school, or secondary school) or school assignment (rural, urban, or suburban). These analyses were completed on the total rating scores for the combined 17 items in the test development and the combined nine items in the use of test scores sections of the questionnaire.

Hypothesis number two was accepted as no significant differences were identified for either the test development or the use of test scores sets of competencies for either the grade level or the school assignment classifications of the respondents. The need rating means for the set of 17 test development competencies for the grade level classification were: elementary 68.86, middle 68.31, and secondary 67.94 (F=0.46, p=.63). The proficiency rating means for this same set of competencies and classification were: elementary 50.45, middle school 49.83, and secondary 49.39 (F=1.07, p=.34). The need means for the set of nine test score use competencies for the grade level classification were: elementary 37.46, middle 37.34, and secondary 36.95 (F=0.51, p=.60). The total proficiency means for this same set of competencies and classification were: elementary 27.55, middle 26.91, and secondary 26.91 (F=1.14, p=.32).

The need means for the set of 17 test development competencies categorized by type of school assignment were: rural 68.37, urban 68.04, and suburban 68.34 (F=0.03, p=.97). The proficiency rating means for this same questionnaire section were: rural 49.84, urban 49.49, and suburban 48.94 (F=0.82, p=.44). The need means for the test score use section of the questionnaire were: rural 36.94, urban 37.10, and suburban 37.42 (F=0.52, p=.60). The proficiency means for this same section were: rural 27.29, urban 26.56, and suburban 26.64 (F=1.40, p=.25).

The lack of significant mean differences among the competency ratings when classified by type of school assignment or by grade level assignment would suggest that the raters were consistent in their ratings of beginning teachers, that beginning teachers were seen as having similar levels of proficiencies despite different grade level or school assignments, and that the principals and supervisors perceived testing and evaluation job needs as being similar for varied school or grade settings.

Hypothesis Three: Testing Versus Other Professional Competencies

In the third section of the questionnaire, the respondents were requested to make an "overall assessment of the preparation of typical beginning teachers in competencies related to tests and evaluation." Three items were provided in this section requiring the principals and supervisors to rate the testing and evaluation competencies of beginning teachers relative to: knowledge of their subject areas, their other professional education competencies, and their overall competencies as educators. Each of these three items had a response scale from one to five, respectively: (1) much below average, (2) somewhat below average, (3) about average, (4) somewhat above average, and (5) well above average.

The responses to this section of the questionnaire were analyzed by various grade levels and types of school assignment for the total group of respondents and by supervisor as compared to principal ratings. When the total group of respondents were

classified by grade assignment of the raters (elementary, middle, or secondary schools) and by type of school(s) the raters were assigned to (rural, urban, or suburban), no significant mean differences within either of the two groups were identified. However, the principals' mean rating as compared to the supervisors' mean rating were significantly different for each of the three items, as indicated in Table 3. The item rating means for each of the three items for these two groups of raters were as follows: knowledge of subject area, principals 3.03, supervisors 2.87 ($t=2.47$, $p=.01$); other professional education competencies, principals 2.96, supervisors 2.81 ($t=2.34$, $p=.02$); and overall competencies as educators, principals 2.93 and supervisors 2.73 ($t=3.34$, $p=.001$). On each item the supervisors' mean rating of the beginning teachers' competencies was lower than the principals' mean rating. The item score means for the total group of respondents (principals plus supervisors) on each of the three items were, respectively, 2.95, 2.89, and 2.84.

Hypothesis three was rejected as item rating means for the principal, supervisor, and the total group of respondents were below average (below 3.0) for eight of the nine rating means. Thus, it is evident that these principals and supervisors perceived beginning teachers as being less competent in testing and evaluation skills as compared to their knowledge and skills in other areas.

Hypothesis Four: Comparison of Principals' and Supervisors' Ratings

A series of independent t tests were used to determine whether or not the supervisors and principals differed significantly in their ratings of beginning teachers' testing and evaluation needs and proficiencies in each of the various competency areas. The results of these analyses for the 17 test development competencies and for the nine test score use competencies are presented on Table 4 and Table 5, respectively.

The comparisons of the principals' ratings to the supervisors' ratings of beginning teacher needs revealed no significant mean difference for the combined 17 test development competencies (principals 68.16 and supervisors 68.63, $t = 0.61$, $p = .54$) or for the combined nine test score use competencies (principals 37.30 and supervisors 37.02, $t = 0.33$, $p = .57$). Further, the comparisons for each individual competency item resulted in the identification of only three significant mean differences ($p < .05$) among the 26 need items from the two sections of the questionnaire. This suggests a high level of agreement between these two groups of raters about the testing and evaluation needs of beginning teachers. Of the three "need" items revealing a significant mean difference between the two groups, two of the items were rated higher by the principals as compared to the supervisors. These two items were: calculating end of term grades, means of 4.04 and 3.87, respectively (Table 5, $t=2.17$, $p=.03$); and deciding the importance of tests and

papers, means of 4.25 and 4.11 respectively (Table 5, t=2.07, p=.04). The third item, use of less formal assessments, was rated as a higher need by the supervisor group: means of 3.70 and 3.54, respectively (Table 4, t=1.98, p=.05).

The series of comparisons between the principals' and supervisors' ratings of beginning teachers' proficiencies revealed a significant mean difference for the combined 17 test development competencies and for the combined nine test score use competencies. These mean differences, respectively, were: principals 50.74 and supervisors 47.81 ($F = 10.91$, $p = .001$), and principals 27.52 and supervisors 26.32 ($F = 5.47$, $p = .02$). Additionally, comparisons for individual competency items resulted in 15 significant mean differences among the 26 items. Each of these identified significant differences revealed a pattern of higher ratings of beginning teachers' testing and evaluation proficiencies by the building principals as compared to the supervisors.

Even though the principals tended to rate the beginning teachers' proficiencies higher than did the supervisors, it is evident from examining the relative item rating magnitudes (ranks) within both sets of proficiency items that the two groups of raters were in rather high agreement about the relative levels of proficiencies. In other words, the principals and supervisors were in high agreement about which proficiencies of the beginning teachers were relatively higher or lower as compared to the total

sets of proficiencies. This was also true of the needs ratings of the two groups of administrators. The Spearman (Rho) correlation coefficients presented on the last lines of Tables 4 and 5 indicate a very high agreement (rank order coefficients of .92 or higher) between the various sets of principals' and supervisors' need or proficiency rating means for both the test development and test score use competencies.

The principals rated beginning teacher proficiencies significantly higher than the supervisors on individual items one through 11 of the 17 test development proficiency items (see Table 4). Similarly, the principals rated the beginning teachers significantly higher on items two, four, and eight (interpreting scores, reteaching needs, and guiding learning) of the nine test score use proficiencies (see Table 5).

In summation relative to hypothesis four, the comparison of principals' and supervisors' ratings resulted in the acceptance of hypothesis four 'a', as the principals and supervisors generally agreed on the relative need for the various beginning teachers' test development and test score use competencies and in the rejection of hypothesis four 'b', as the principals and supervisors significantly differed in their ratings of beginning teachers' proficiencies in both sections of the questionnaire. This data indicate that the principals rated beginning teachers as having higher test development and test score use proficiencies than did

the supervisors, but the two groups of raters were in general agreement about the relative (separate rank orders) proficiency levels of the beginning teachers within the set of 26 competencies.

Summary and Discussion

The analyses of the data collected resulted in the rejection of hypothesis one as the combined supervisors and principals' ratings of the beginning teachers' needs were significantly higher than their ratings of beginning teachers' proficiencies for each of the 26 test and evaluation competency areas. This would suggest that the total group of respondents viewed beginning teachers' proficiencies in the area of tests and evaluation to be less than adequate in terms of typical job needs. It might suggest also that those professionals responsible for inservice and preservice teacher training should give more attention to testing and evaluation skills development.

Hypothesis two was accepted as the combined supervisor and principal respondents grouped by different types of schools (rural, urban, or suburban) or by grade level assignments (elementary, middle, or secondary grades) did not significantly vary in either their ratings of teachers' testing and evaluation competency needs or their ratings of beginning teachers' testing and evaluation proficiencies. This would suggest that testing and evaluation needs or proficiencies do not vary greatly from grade to grade or school to school and that inservice training sessions

might include various school and grade level personnel without being detrimental to the learning process.

Hypothesis three was rejected as the combined principal and supervisor respondents did not rate beginning teachers' test and evaluation skills to be as high or higher than their knowledge of subject areas, as high or higher than their other professional education competencies, or as high or higher than their overall competencies as educators. This might suggest that beginning teachers' testing and evaluation skills are less well developed than their other professional skills, and it further might confirm that preservice and inservice trainers of teachers ought to give more attention to testing and evaluation skill development.

Lastly, hypothesis four 'a' was accepted but four 'b' was rejected. The separate principals' and supervisors' ratings revealed a high degree of agreement between these two groups in rating the needs of beginning teachers for the various test development and test score use competencies. However, these two groups of raters differed significantly in their ratings of beginning teacher proficiencies in the various test development and test score use competencies. Generally, the principals rated the proficiencies of beginning teachers higher than did the supervisors. Whether this difference in rating levels is a consequence of differences in opportunities to observe beginning teachers or of differences in relative rating tendencies of the

two groups of respondents could not be determined from the data. The latter might be considered less likely, however, as the two groups rated needs in very similar manner.

Overall, the data collected suggested considerable agreement among the principals and supervisors in their relative ratings of typical teachers' needs and beginning teachers' proficiencies on the 26 identified testing and evaluation competency areas. The relative mean rating magnitudes within each group of competencies for both the principals and supervisors were very similar (Spearman Rho's of .92 and higher). This consistency along with rating stability found across respondent grade and school assignments, and the very high consistency in principals' and supervisors' mean ratings of needs would appear to encourage one's confidence in this data.

The magnitude of the discrepancy scores (need mean minus proficiency mean) for each of the 26 testing competencies should provide those concerned about either inservice or preservice teacher training with a practical guide in designing content for such training programs (true, this sample was limited to a single state, but it is a populous state which employs beginning teachers trained in many other states). It would appear that teacher trainers might wish first to address competencies associated with highly rated needs but with large discrepancy scores. Rather specifically, this set of data would suggest that inservice or

preservice training might best emphasize the use of tests and scores for reteaching of content, guiding student learning, and positively influencing study and learning. Further, this data would suggest that more practice be given in skills such as stating measurable objectives, writing items that assess true student progress, writing items that measure higher thinking processes, and in the writing and scoring of essay items.

References

Balch, J. (1964). The influence of the evaluating instrument on students' learning. American Educational Research Journal, 6, 169-182.

Coffman, W. E. (1971). Essay examinations. In R. L. Thorndike (Ed.), Educational measurement (2nd ed., pp. 271-302). Washington, D.C.: American Council on Education.

Dwyer, C. A. (1982). Achievement testing. In H. E. Mitzel (Ed.), Encyclopedia of educational research (4th ed., Vol. 1, pp. 13-22). New York: The Free Press.

D'Ydewalle, G., Swerts, A., & DeCorte, E. (1983). Study time and test performance as a function of test expectations. Contemporary Educational Psychology, 8, 55-67.

Fleming, M. & Chambers, B. (1983). Teacher-made tests: Windows on the classroom. New directions for testing and measurement, 19, 29-38.

Green, R. L. (1975). Tips on educational testing: What teachers and parents should know. Phi Delta Kappan, October, 89-93.

Gullickson, A. R. (1984). Teacher perspectives of their instructional use of tests. Journal of Educational Research, 77, 244-248.

Hanna, G. S. (1976). Effects of total and partial feedback in multiple-choice testing upon learning. Journal of Educational Research, 69, 202-205.

Kulhavy, R. W., Dyer, J. W., & Silver, L. (1975). The effects of notetaking and test expectancy on the learning of text material. Journal of Educational Research, 68, 363-365.

Lambert, B. R. (1980-81). Teacher attitudes on testing: A multiple perspective. College Board Review, Winter, 13-14 and 29-30.

Madaus, G. F. (1985). Public policy and the testing profession: You've never had it so good? Educational Measurement Issues and Practice, 4, 5-11.

Nungester, R. J., & Duchastel, P. C. (1982). Testing versus review: Effects on retention. Journal of Educational Psychology, 74, 18-22.

Peckham, P. D., & Roe, M. D. (1977). The effects of frequent testing. Journal of Research and Development in Education, 10, 40-50.

Rogers, B. G. (1985). Prospective teacher perceptions of how classroom teachers use evaluation methods: A qualitative research approach. Mid-Western Educational Research, 6, 13-20.

Shaha, S. (1984). Matching-tests: Reduced anxiety and increased test effectiveness. Educational and Psychological Measurement, 44, 869-881.

Stewart, L. G., & White, M. A. (1976). Teacher comments, letter grades, and student performance: What do we really know? Journal of Educational Psychology, 68, 488-500.

Wexley, K. N., & Thornton, C. L. (1972). Effect of verbal feedback of test results upon learning. Journal of Educational Research, 66, 119-121.

1/A

4/16/87

Table 1

Principals and Supervisors' Estimates of the Needs and Proficiencies of Beginning Teachers in 17 Test Development Competency Areas

	Test Development Competencies	Need	Proficiency	Discrepancy	Rank*	t	p
1.	Writing multiple choice items	3.83	2.99	.84	12	19.53	.001
2.	Writing completion items	3.91	3.06	.85	11	19.75	.001
3.	Writing matching items	3.70	3.10	.60	15	13.73	.001
4.	Writing true/false items	3.51	2.99	.62	14	10.68	.001
5a.	Writing essay items	4.27	2.74	1.53	5.5	32.29	.001
5b.	Scoring essay items	4.35	2.67	1.68	3	36.06	.001
6.	Identifying good and poor items	4.34	2.83	1.51	7	35.15	.001
7.	Items harmony school/class goals	4.33	2.79	1.54	4	34.12	.001
8.	Stating clear/measurable objectives	4.40	2.87	1.53	5.5	33.26	.001
9.	Items measure higher thinking	4.45	2.55	1.90	1	38.29	.001
10.	Items measure true progress	4.50	2.78	1.72	2	38.39	.001
11.	Use less formal assessments	3.61	2.86	.75	13	15.95	.001
12.	Use observation assessments	4.02	2.96	1.06	9.5	24.14	.001
13.	Use sociometric type assessments	3.19	2.72	.47	16.5	10.70	.001
14.	Selecting items from manuals	3.60	3.13	.47	16.5	11.24	.001
15.	Attractive test format	4.08	3.02	1.06	9.5	24.46	.001
16.	Test coverage of text and class	4.51	3.19	1.32	8	32.18	.001
	Combined items totals	68.68	49.23				
	t-ratio		38.70				
	Probability level		.001				

*Rank ordered by magnitude of discrepancy

Table 2

Principals and Supervisors' Estimates of the Needs and Proficiences of Beginning Teachers in Nine Test Score Use Competency Areas

	Test Score Use Competencies	Need	Proficiency	Discrepancy	Rank*	t	p
1.	Calculation, means, SD's, reliability	3.04	2.42	.62	9	12.97	.001
2.	Interpreting scores and student progress	4.24	2.88	1.36	5	31.65	.001
3.	Identifying individual/class strength/weakness	4.33	2.95	1.38	4	35.27	.001
4.	Determining reteaching needs	4.55	2.88	1.67	1	36.79	.001
5.	Use of tests and grades to influence learning	4.31	2.86	1.45	3	31.75	.001
6.	Calculating end of term grades	3.98	3.34	.64	8	15.46	.001
7.	Grading tests, papers, etc.	4.12	3.41	.71	7	17.86	.001
8.	Deciding importance tests, papers, etc.	4.20	3.15	1.05	6	25.05	.001
9.	Deriving information tests/guide learning	4.38	2.90	1.48	2	33.80	.001
	Combined items totals	37.28	26.77				
	t-ratio	37.84					
	Probability level	0.000					

*Rank ordered by magnitude of discrepancy

Table 3

Beginning Teachers' Testing Proficiencies Compared to Their Other Proficiences, As Estimated by Principals and Supervisors

Relative Proficiency Rating Items*	Principal	Supervisor	Total	t**	p
1. Relative to knowledge of their subject areas, beginning teachers' test and evaluation competencies are...	3.03	2.87	2.95	2.47	.014
2. Relative to their other professional education competencies, such as planning, discipline, etc., beginning teachers' test and evaluation competencies are....	2.96	2.81	2.89	2.34	.020
3. Relative to their overall competencies as educators, beginning teachers' test and evaluation competencies are...	2.93	2.73	2.84	3.34	.001

*Ratings were recorded via a five point Likert-type scale, 5 (well above average), 4 (somewhat above average), 3 (above average), 2 (somewhat below average), and 1 (much below average)

**Mean comparisons between principal and supervisor ratings

Table 4

Beginning Teachers' Test Development Competency Need and Proficiency Means, As Rated by Supervisors and Principals

		Need Rating Means				Proficiency Rating Means			
	Test Development Competencies	Prin.	Super.	t	p	Prin.	Super.	t	p
1.	Writing multiple choice items	3.79	3.89	1.32	.198	3.06	2.91	2.67	.008
2.	Writing completion items	3.90	3.89	0.24	.810	3.13	2.97	2.60	.010
3.	Writing matching items	3.73	3.65	1.01	.312	3.16	3.04	2.07	.039
4.	Writing true/false items	3.55	3.45	1.11	.267	3.06	2.90	2.50	.013
5a.	Writing essay items	4.20	4.32	1.66	.097	2.85	2.59	3.69	.001
5b.	Scoring essay items	4.30	4.40	1.37	.171	2.78	2.53	3.42	.001
6.	Identifying good and poor items	4.30	4.35	0.78	.436	2.92	2.73	2.98	.003
7.	Items harmony school/class goals	4.30	4.35	0.76	.448	2.88	2.72	2.19	.029
8.	Stating clear/measurable objectives	4.35	4.42	1.06	.288	2.97	2.73	3.34	.001
9.	Items measure higher thinking	4.39	4.51	1.85	.065	2.65	2.43	2.91	.004
10.	Items measure true progress	4.47	4.51	0.55	.581	2.88	2.65	3.27	.001
11.	Use less formal assessments	3.54	3.70	1.98	.048	2.93	2.79	2.15	.032
12.	Use observation assessments	3.96	4.08	1.79	.075	3.02	2.91	1.62	.106
13.	Use sociometric type assessments	3.22	3.13	1.26	.207	2.73	2.76	0.41	.680
14.	Selecting items from manuals	3.59	3.57	0.29	.773	3.16	3.12	0.66	.511
15.	Attractive test format	4.06	4.08	0.21	.834	3.05	3.01	0.64	.523
16.	Test coverage of text and class	4.49	4.53	0.66	.507	3.24	3.14	1.53	.127
	Combined items total	68.16	68.63			50.48	47.88		
	t-ratio		0.61				3.34		
	Probability level		0.54				.001		
	Spearman Rho's Between Ranks of Means		0.98				0.92		

Table 5

Beginning Teachers' Test Score Use Competency Need and Proficience Means, As Rated by Supervisors and by Principals

Test Score Use Competencies	Need Rating Means				Proficiency Rating Means			
	Prin.	Super.	t	p	Prin.	Super.	t	p
1. Calculation, means, SD's, reliability	3.02	3.04	0.24	.813	2.44	2.38	0.75	.455
2. Interpreting scores and student progress	4.20	4.27	0.96	.337	2.95	2.77	2.51	.012
3. Identifying individual/class strength/weakness	4.39	4.48	1.42	.156	3.01	2.88	1.85	.065
4. Determining reteaching needs	4.53	4.59	0.97	.335	2.96	2.80	2.06	.040
5. Use of tests and grades to influence learning	4.31	4.32	0.16	.875	2.92	2.79	1.90	.058
6. Calculating end of term grades	4.04	3.87	2.17	.031	3.40	3.30	1.50	.134
7. Grading tests, papers, etc.	4.13	4.06	0.96	.336	3.47	3.36	1.72	.085
8. Deciding importance tests, papers, etc.	4.25	4.11	2.07	.039	3.26	3.04	3.44	.001
9. Deriving information tests/guide learning	4.38	4.38	0.09	.930	2.99	2.83	2.29	.022
Combined items total	37.30	37.02			27.52	26.32		
t-ratio		0.33				5.47		
Probability value		.569				.020		
Spearman Rho's Between Ranks of Means		0.98				0.98		